

Claims

1. A method for rerouting network traffic, operating in conjunction with a server cluster comprising at least a first and a second server, the method comprising the steps of:
 - receiving a packet in a first server, the packet containing indication of a source address;
 - adding a redirection header to the packet, the redirection header including a service address belonging to the second server; and,
 - transmitting the packet and redirection header to the service address.
2. The method of claim 1 further comprising the step of obtaining the service address from a redirection database.
3. The method of claim 1, wherein the source address and the service address are composed of an Internet address and a port.
4. The method of claim 1, wherein the redirection header further comprises a redirection flag indicating whether a client having the source address is expected to perform a redirection to the service address.
5. The method of claim 4, wherein the redirection header further includes a permanent flag indicating whether the client is expected to perform the redirection permanently to the service address.
6. The method of claim 4, wherein the redirection header further includes a once flag indicating whether the client is expected to perform the redirection once to the service address.
7. The method of claim 4, wherein the redirection header further includes a leased flag indicating whether the client is expected to perform the redirection to the service address as long as a lease is in force.
8. The method of claim 1 further comprising the step of receiving, by a server in the cluster, an indication of redirection failure; and,

providing a second redirection header to the packet, the redirection header including a service address belonging to an alternative third server.

9. The method of claim 1 further comprising the steps of:
for each server of the server cluster:

responsive to transmission of a UDP packet, creating and storing a record comprising information identifying the sender address and sender port, and recipient address and recipient port of the UDP packet, and information reflecting the time of transmission;
responsive to additional transmission of UDP packet from same sender address and port to same recipient address and port, updating said time information to reflect the time of said additional transmission;
periodically checking for said time information; and,
removing said record if a predetermined period elapsed from said time of transmission to the time of said checking.

10. The method of claim 9 further comprising the steps of:

composing an artificial UDP session related to each record comprising information identifying the sender address,
counting periodically artificial UDP sessions resulting in the number of UDP sessions,
counting or obtaining TCP sessions resulting in the number of UDP sessions,
obtaining the total load of each server of the server cluster as a function of the UDP sessions and TCP sessions.

11. The method of claim 10 further comprising the steps of:

using the total load of each server when performing the load balancing of the server cluster.

12. The method of claim 10 further comprising the steps of:

using the total load of each server when controlling the usage of the communication links used by the server cluster.

13. The method of claim 2, wherein said redirection database further collects information regarding load on a selected server set, and wherein said information is used as criteria for selecting second server.
14. A method for redirecting network traffic comprising the steps of:
 - receiving a service request in a server, said request having indication of the request sender;
 - selecting a second server for servicing the service request; and,
 - sending to said sender, a redirection indication to said second server.
15. A server cluster having a plurality of member servers, and adapted to reroute network traffic, the cluster comprising:
 - a redirection database adapted to receive information indicative of a client request, and responsively provide a service address comprising an address of a target server capable of servicing said client request; and,
 - a receiving server adapted to receive said client request and send information indicative thereof to said database;
 - a transmitting server adapted to send said service address to the client that originating the request.
16. The server cluster according to claim 15, wherein said receiving server and said transmitting server are the same server.
17. The server cluster of claim 15 wherein the redirection database is integrated in said transmitting server or receiving server or a server combining both.
18. The server cluster of claim 15 wherein the target server is selected according to criteria comprising consideration of server load of a plurality of servers in the cluster.
19. The server cluster of claim 15, wherein a plurality of said member servers are adapted to act as said receiving server.
20. A redirecting client adapted for network traffic rerouting, said client comprising:
 - a first module adapted to send a service request to a first server;
 - and,

a second module to resend the service request to a second server, responsive to receiving a communication packet comprising a redirection header having a service address containing the address of said second server, or an indication thereof.

21. The client of claim 20 further comprising a module adapted to resend said service request to said first server, responsive to service failure of said second server;

wherein said resent service request contains indication of said service failure.

22. The client of claim 20, wherein said second module is further adapted to direct subsequent communications relating to said service request to said second server.
23. The client of claim 20 further adapted to send consequent service requests to said second server responsive to indication of doing so embedded within said redirection header.
24. The client of claim 20 wherein said first module is further adapted to indicate a redirecting capacity in said service request.
25. The client of claim 24 wherein said indication of redirecting capacity is indicated in a manner that allows a server lacking redirection capability to service said service request.